Predicting cycle hires

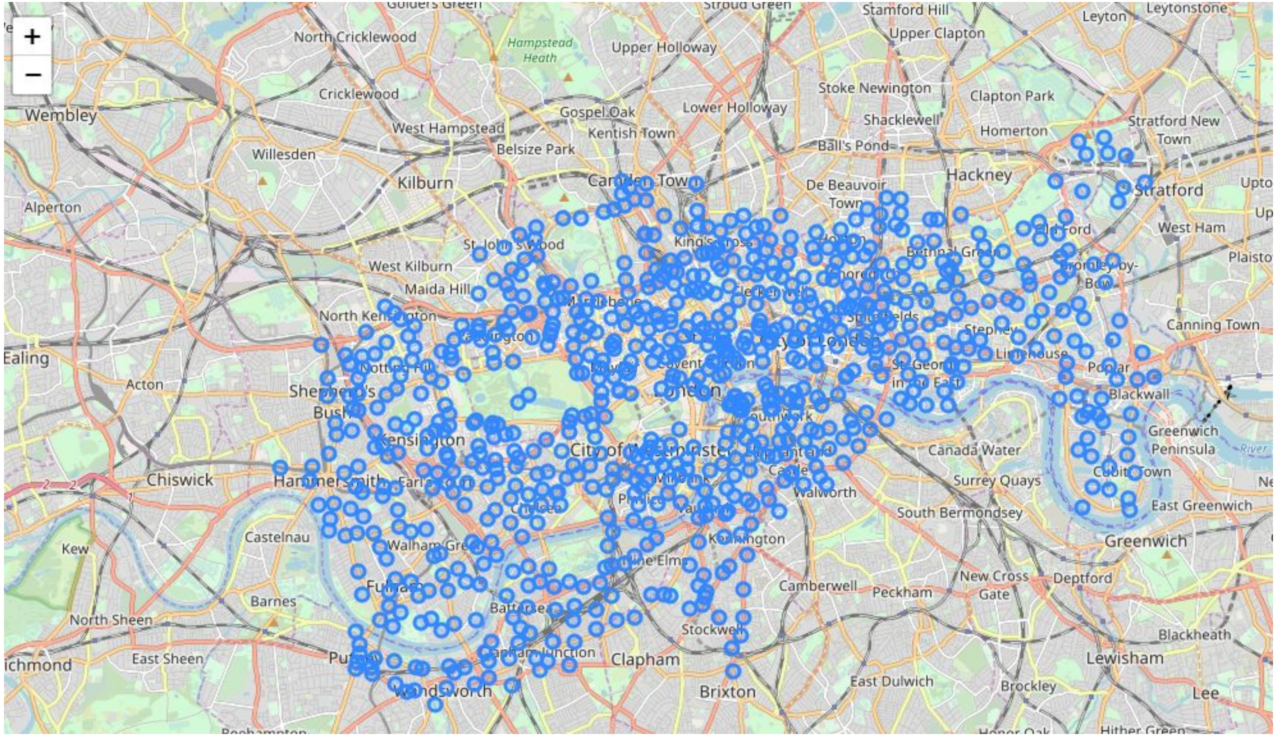
Using TFL(Transport for London) data

Project Proposal

1. Understanding the current utilization of the public hire scheme using historical data

2. Predicting the future demand for bikes daily at each station to improve availability of bikes and optimize flow of system

Location of cycle docks



Feature Engineering

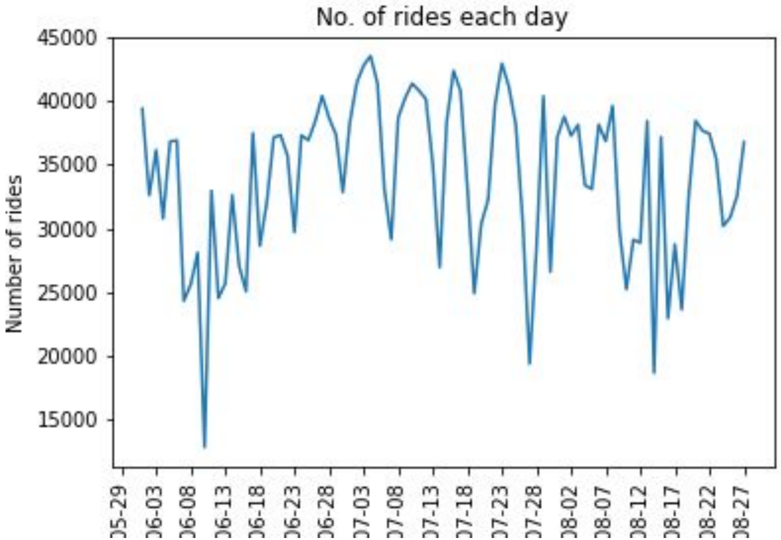
Data provided by TFL:

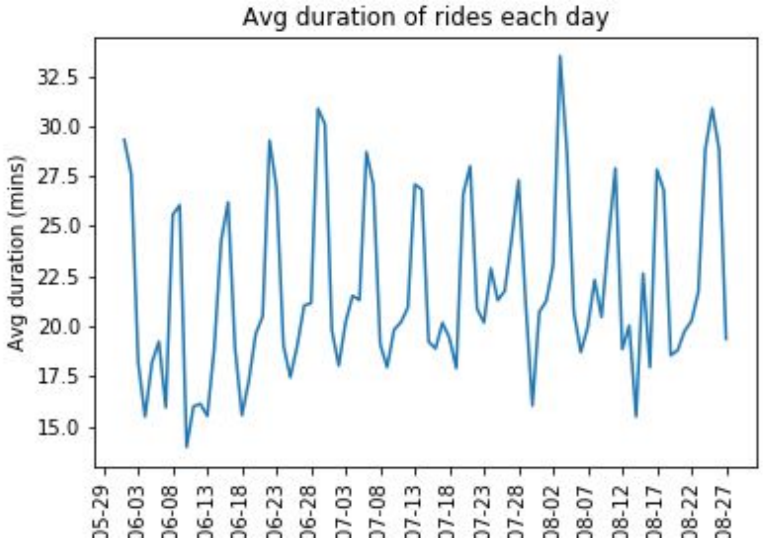
- Start time, End time, Duration
- Start Station ID, End Station ID

Data from weather API:

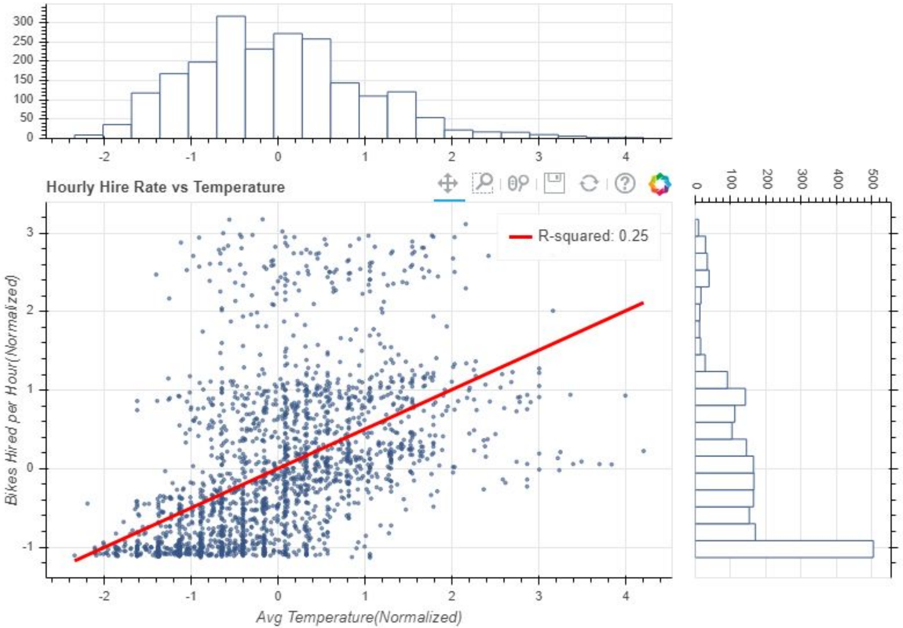
- Temperature
- Wind speed
- Weather Condition

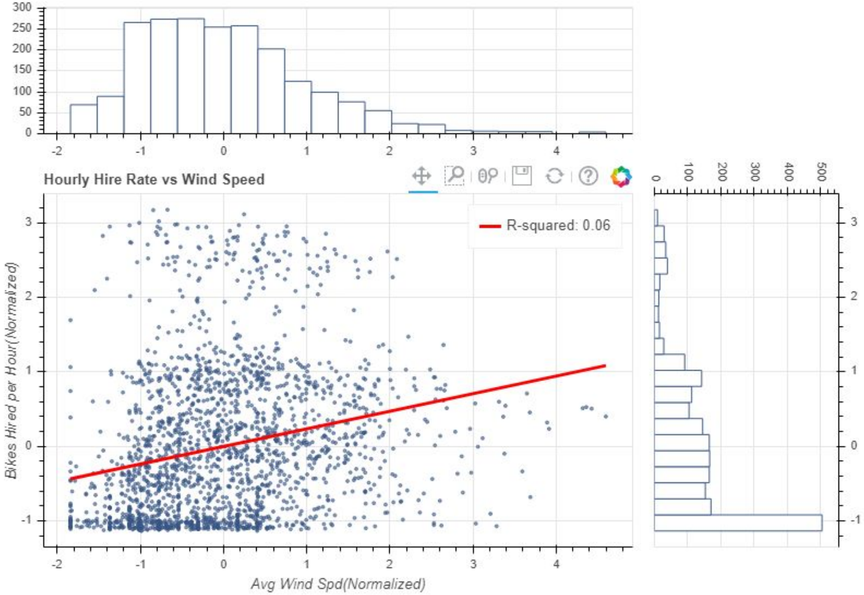
Data Visualization



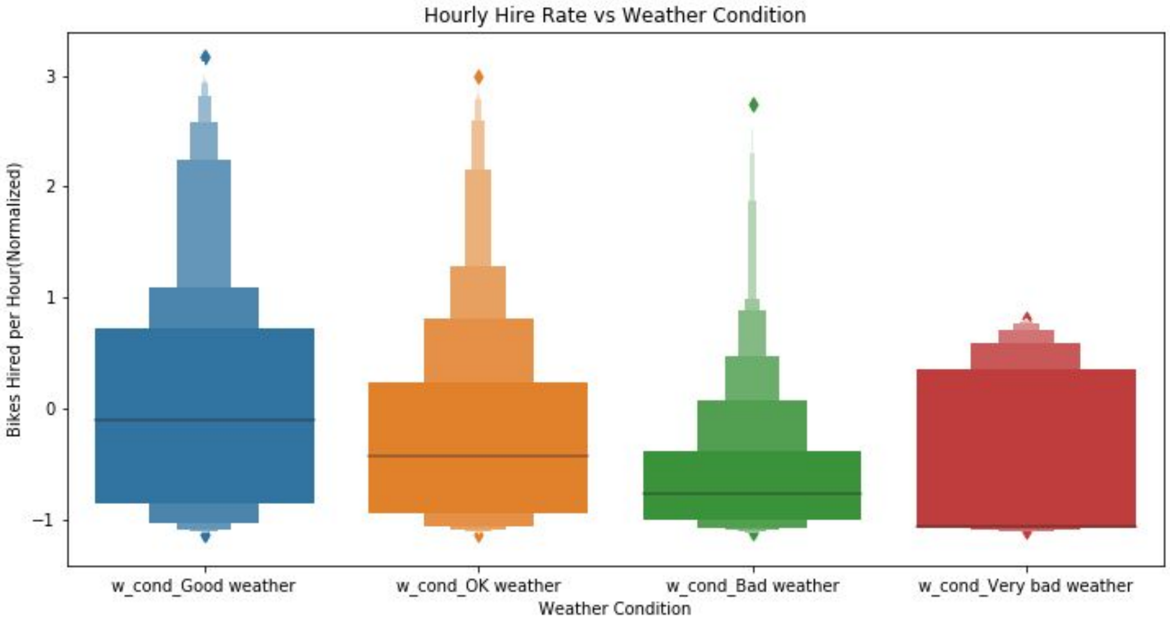


Relationship between weather and hiring frequency

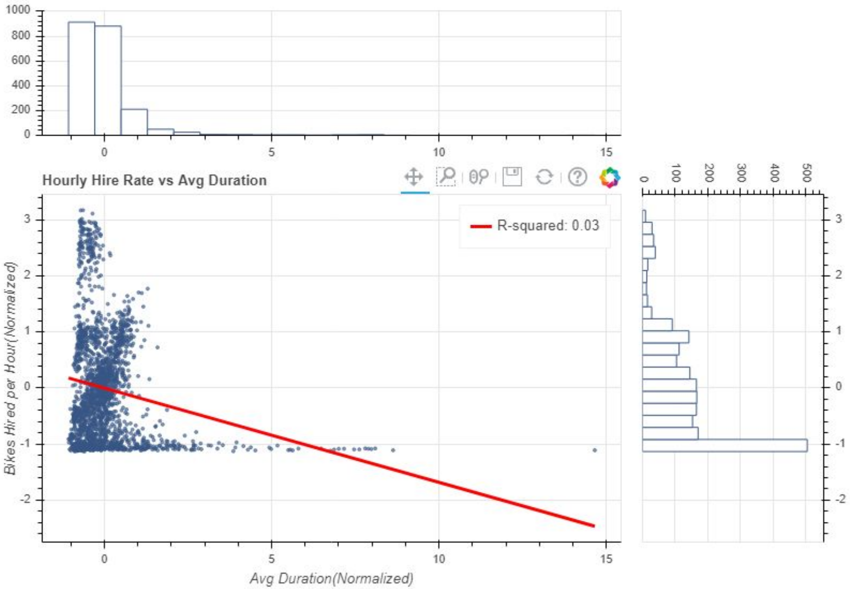


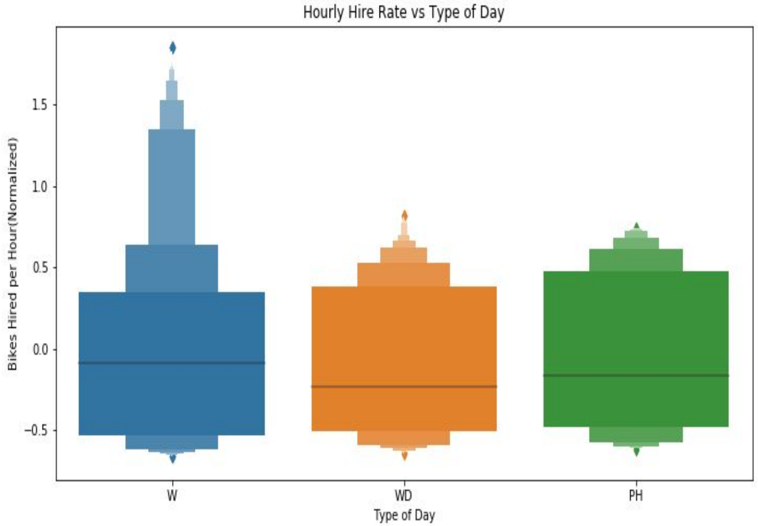


Relationship between weather and hiring frequency



Exploring more relationships





Final Dataset

Features:

- Day of week
- Is weekday?
- Number of hires from station day before
- Number of bikes docked at station day before
- 7 day rolling duration
- Temperature
- Wind speed
- Good weather, OK weather, Bad weather, Very bad weather

Target

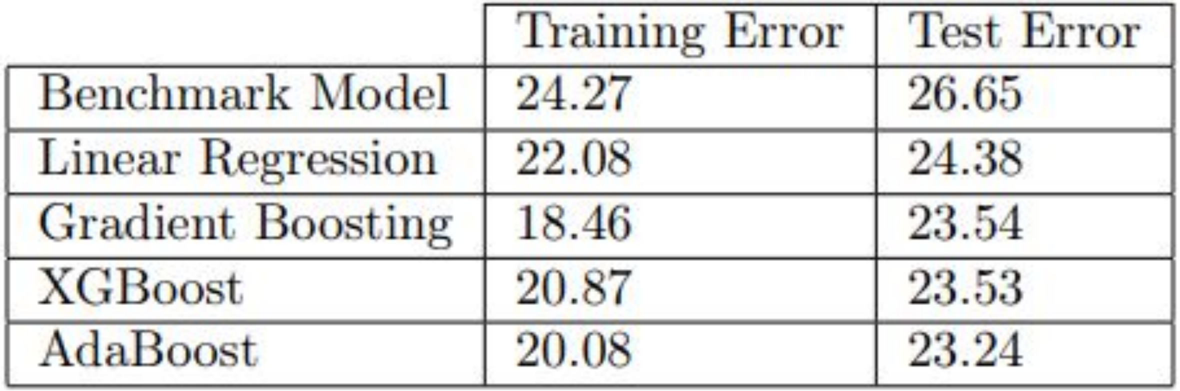
- Number of cycles hired on a future date

Models

- Benchmark Model- assume number of hires on a given day = number of hires day before
- Linear Regression, Ridge Regression
- Gradient Boosting- implemented in scikit-learn
- XGBoost
- AdaBoost

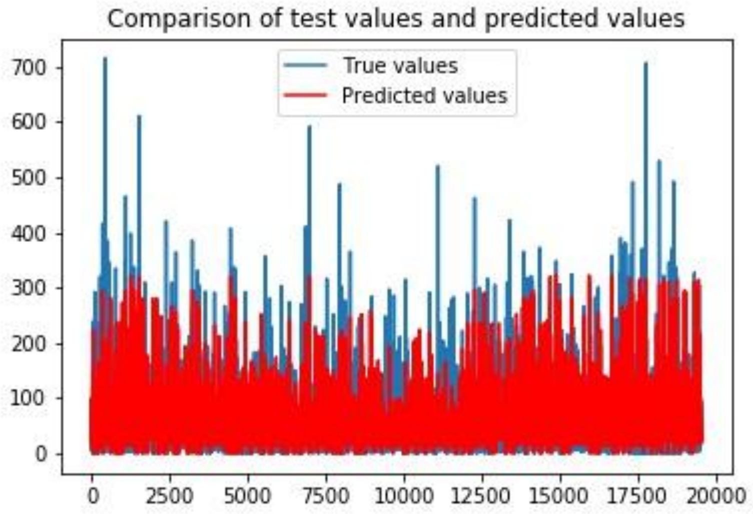
Results

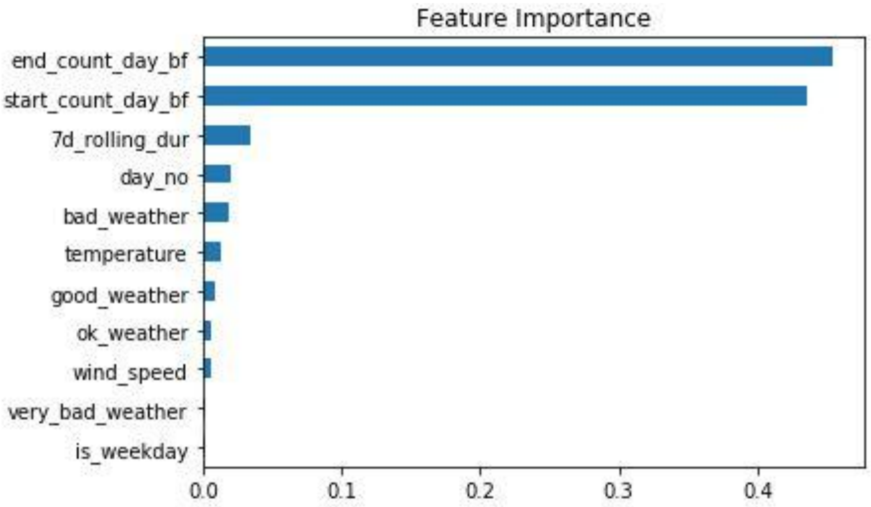
- Using RMSE



	Training Error	Test Error
Benchmark Model	24.27	26.65
Linear Regression	22.08	24.38
Gradient Boosting	18.46	23.54
XGBoost	20.87	23.53
AdaBoost	20.08	23.24

AdaBoost results





Analysis

- AdaBoost is winner but not significant improvement from Linear Regression
- Models unable to do peak prediction well- some stations are very popular and peaks come from the same 3 stations
- Very high feature importance on day before hires and docks

Future Work

- Add locational information such that model can learn the idea of different stations
- Separate training of dataset by popular stations and less popular stations
- Other possible features: Information of tube disruptions